

# DRAFT

## On the use of an Arbitrary Lagrangian-Eulerian vertical coordinate in global atmospheric modeling

Rainer Bleck\*, Stan Benjamin, Jin Lee, and Alexander E. MacDonald  
NOAA Earth Systems Research Laboratory  
Boulder, Colorado

March 20, 2009

### Abstract

This is one of several articles describing the functionality of the global weather prediction model FIM (for Flow-following, Finite-volume, Icosahedral) developed at NOAA's Earth Systems Research Laboratory. Emphasis in the present article is on the design of the vertical coordinate – the “flow-following” aspect of FIM. The coordinate is terrain-following near the ground and isentropic in the free atmosphere. The spatial transition between the two coordinates is adaptive and is based on the ALE (Arbitrary Lagrangian Eulerian) paradigm. The impact of vertical resolution tradeoffs between the present hybrid approach and traditional terrain-following coordinates is demonstrated in a three-part case study.

tent to which the magnitude of individual terms in a differential equation is misrepresented in a numerical model due to grid resolution limits. There are at least three ways to reduce the truncation error: one can (1) refine the grid, (2) approximate spatial and temporal derivatives in a more precise way, or (3) optimize the placement of grid points in the space-time domain.

FIM, the flow-following, finite-volume, icosahedral NWP model recently developed by NOAA's Earth Systems Research Laboratory, takes advantage of recently developed ideas about grid point placement – the third alternative just mentioned. Specifically, FIM uses the icosahedron, a near-spherical body composed of 20 equilateral triangles, as a basis for horizontal grid layout while in the vertical it uses a coordinate with a strong Lagrangian (hence “flow-following”) flavor. The focus of the present article is on the vertical coordinate in FIM.

### 1. Introduction

The last few decades have seen significant advances in numerical weather prediction. The skill of today's NWP models owes much to improved closure schemes for physical processes that are too short-lived or too small in scale to be resolved by a model's space-time mesh. Higher numerical accuracy, made possible primarily by faster computers but to some degree by the invention of new techniques for solving partial differential equations, also had a large impact on forecast skill.

Numerical accuracy is typically expressed in terms of the truncation or discretization error, defined as the ex-

### 2. History

Physical reasoning suggests that entropy or a variable related to it, such as buoyancy, is the most appropriate candidate for a Lagrangian vertical coordinate in modeling stratified, quasi-adiabatic flow. The advantages of analyzing atmospheric motion in the free atmosphere in an isentropic reference frame have been thoroughly discussed in the literature starting with Rossby and collaborators (1937), Rossby (1940), and Kleinschmidt (1950). Those arguments need not be repeated here. Likewise, the rationale for using isentropic coordinates in NWP models has been laid out repeatedly [e.g., Eliassen and

---

\*Primary affiliation: NASA Goddard Institute for Space Studies, Columbia Univ., New York

Raustein (1968), Bleck (1974), Hsu and Arakawa (1990), Benjamin et al. (2004)]. The list of potential advantages of isentropic modeling compiled by those and other authors is long, and there is not much we can add to it at this time.

Isentropic NWP models came into being in the 1960s, but it is fair to say that “pure” isentropic coordinate models (those that use entropy as vertical coordinate throughout the model domain) have not withstood the test of time because of their inherent inability to provide vertical resolution in unstably stratified air columns. Complexities associated with coordinate-ground intersections also were, and continue to be, a deterrent to pure isentropic coordinate modeling.

Early experiments with isentropic models were conducted to simulate baroclinic instability with an eye on short-range weather prediction. Since modeling of diabatic forcing was not essential in that context, the intersection of coordinate surfaces with the ground – unavoidable in baroclinic flow – was regarded as the main numerical challenge. Eliassen and Raustein (1968), in their pioneering work on primitive-equation isentropic modeling, chose to track coordinate-ground intersections by solving an advection equation for surface potential temperature  $\theta_s$ ,

$$\frac{\partial \theta_s}{\partial t} + \mathbf{v}_s \cdot \nabla \theta_s = 0. \quad (1)$$

Though justifiable at the time, this strategy created a redundancy problem because isentropic coordinate models typically contain another equation predicting the location of coordinate-ground intersections, namely, the continuity equation for isentropic layer thickness  $\Delta p$ ,

$$\frac{\partial \Delta p}{\partial t} + \nabla \cdot (\mathbf{v} \Delta p) = 0. \quad (2)$$

To understand why this is so, keep in mind that the line marking the intersection of an isentrope with the ground also marks the edge of the region where the thickness of the coordinate layer beneath the isentrope is zero. In principle, solving (2) in the vicinity of the intersection line therefore provides information about which way the line is moving.

Obtaining accurate solutions of (2) in the transition region between zero and nonzero  $\Delta p$  values is numerically challenging. Since the solution of (1) is subject to numerical errors as well, the two equations do not always agree on where a coordinate surface intersects the ground at any given time. The resulting discrepancies act at best

as a source of model noise; at worst, they lead to numerical instability.

The problem just described spawned several attempts in the 1970s to improve the treatment of coordinate-ground intersections by inserting a set of terrain-following coordinate surfaces, commonly referred to as  $\sigma$  surfaces (Phillips 1957), between the ground and the isentropic domain (Deaven 1976; Uccellini et al. 1979; Friend et al. 1977; Bleck 1978). In all these schemes, with the exception of scheme D of Bleck (1978), the two coordinate domains overlap, i.e., coordinate surfaces belonging to one domain intersect those belonging to the other. This requires interpolation. Scheme D, having no overlaps, leads to a particularly simple set of model equations, making it easy to formulate them in rigorously conservative form, but it has shortcomings of its own. The Uccellini et al. (1979) scheme and a variant of Bleck’s scheme D are in use today in different versions of the University of Wisconsin global model [Zapotocny et al. (1994), Schaak et al. (2004)].

The decade of the 1980s saw some progress in the related field of ocean circulation modeling with an entropy-related vertical coordinate. Specifically, Bleck and Boudra (1981) developed a coordinate system which is mainly isopycnic but allows coordinate layers to turn into constant-thickness layers beneath the sea surface to overcome the massless layer problem associated with modeling baroclinic ocean states. This may have been the first time that ALE (Arbitrary Lagrangian-Eulerian) coordinates were used in geophysical modeling.

Increasingly skillful schemes for avoiding coordinate-ground intersections were developed in the 1990s [Bleck and Benjamin (1993), Konor and Arakawa (1997), Webster et al. (1999)]. These more recent schemes resemble scheme D of Bleck (1978) in the sense that they avoid overlaps between isentropic and terrain-following or isobaric subdomains, but they go beyond scheme D in alleviating, to varying degrees, the abrupt transition from non-isentropic to isentropic coordinate representation. Some aspects of these schemes will be discussed later in greater detail.

A parallel effort to build purely isentropic or isopycnic models free of the redundancy implied by jointly solving (2) and (1) was brought to a conclusion [Bleck (1984), Bleck and Boudra (1986)] with the advent of the Flux Corrected Transport algorithm (Boris and Book 1973). This scheme yielded well-behaved solutions of the con-

tinuity equation in transition zones between zero and nonzero layer thickness where under- and overshoots (especially negative  $\Delta p$  values) are particularly detrimental to numerical stability. An alternative approach to pure isentropic modeling, advocated by Hsu and Arakawa (1990), is to compute horizontal mass fluxes using the Takacs (1985) advection scheme. Since this scheme does not enforce positive definiteness as rigorously as does FCT, Hsu and Arakawa (1990) had to keep a small amount of mass in coordinate layers that in the aforementioned models are allowed to become truly massless.

Having assembled a set of numerically resilient tools for handling the intersection of isentropes with the ground, the isentropic modeling community turned its attention to the second problem in isentropic modeling, i.e., the need to accomodate unstable lapse rates associated with diabatic surface forcing. Since both problems manifest themselves at or near the surface, the remedies developed to address the coordinate-ground intersection problem also were useful in alleviating the unstable lapse rate problem. Models general enough to accomodate diabatic surface forcing [Zhu et al. (1992), Bleck and Benjamin (1993), Zapotocny et al. (1994), Konor and Arakawa (1997), Schaak et al. (2004)] therefore resemble the ones mentioned earlier in their vertical layout. The salient aspect of all these approaches is that they “hybridize” the isentropic coordinate, i.e., strike a compromise between Lagrangian and Eulerian vertical grid representation.

### 3. Brief review of FIM dynamics

The horizontal aspects of the dynamical core of FIM are described in detail elsewhere (Lee and MacDonald 2009; Lee et al. 2009), but a brief review is needed here to put the ensuing discussion of the vertical coordinate into context [see also Kasahara (1974) and Bleck (1978) for derivations of equations in generalized vertical coordinates].

FIM solves a system of layer-integrated, hydrostatically approximated conservation equations for momentum, mass, thermal energy, and gaseous as well as liquid or frozen water content. “Physics” components that model diabatic forcing of the atmosphere – such as water phase changes, radiation, and surface fluxes – were

imported from the Global Forecast System of NOAA’s Centers for Environmental Prediction. Details about their implementation will be the topic of a forthcoming paper on FIM’s performance in general.

In the following, let  $\mathbf{v}$  be the (Cartesian) horizontal velocity vector taken as vertically constant within a coordinate layer;  $\nabla_s$  the 2-D gradient operator at  $s = \text{const}$ ;  $\Pi = c_p(p/p_0)^{R/c_p}$  the Exner function;  $\theta = c_p T/\Pi$  the potential temperature;  $M = gz + \Pi\theta$  the Montgomery potential;  $\zeta$  the vorticity (i.e., the vertical or  $\mathbf{k}$  component of the velocity curl vector);  $\dot{\theta}$  the net diabatic heating; and  $\mathbf{F}$  the sum of frictional forces.

The layer-integrated conservation equations, supplemented by the hydrostatic equation, are

*Momentum conservation:*

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} + (\zeta + f)\mathbf{k} \times \mathbf{v} \\ + \frac{1}{\Delta p} \left[ \left( \dot{s} \frac{\partial p}{\partial s} \right)_2 (\hat{\mathbf{v}}_2 - \mathbf{v}) - \left( \dot{s} \frac{\partial p}{\partial s} \right)_1 (\hat{\mathbf{v}}_1 - \mathbf{v}) \right] \\ + \nabla_s \left( M + \frac{\mathbf{v}^2}{2} \right) - \Pi \nabla_s \theta = \mathbf{F}. \end{aligned} \quad (3)$$

Here, indices 1 and 2 denote the upper and lower interface, respectively, of a coordinate layer, and  $\Delta p = p_2 - p_1$ . The vertical advection terms (those involving  $\dot{s}$ ) involve interface velocity values denoted here by  $\hat{\mathbf{v}}$ . Since  $\mathbf{v}$  is discontinuous at interfaces, the definition of  $\hat{\mathbf{v}}$  is to some extent arbitrary and in practice depends on the finite-difference vertical advection scheme.

*Mass conservation:*

$$\frac{\partial \Delta p}{\partial t} + \nabla_s \cdot (\mathbf{v} \Delta p) + \left( \dot{s} \frac{\partial p}{\partial s} \right)_2 - \left( \dot{s} \frac{\partial p}{\partial s} \right)_1 = 0. \quad (4)$$

*Thermal energy conservation:*

$$\frac{\partial (\theta \Delta p)}{\partial t} + \nabla_s \cdot (\theta \mathbf{v} \Delta p) + \left( \dot{s} \frac{\partial p}{\partial s} \hat{\theta} \right)_2 - \left( \dot{s} \frac{\partial p}{\partial s} \hat{\theta} \right)_1 = \dot{\theta} \Delta p. \quad (5)$$

As before, the caret in (5) denotes interface values needed in the vertical transport terms. The method by which they are constructed (upstream, centered,...) determines properties such as monotonicity and diffusiveness of vertical advection in the model.

Equations for other mass field tracers (moisture etc.) have the same form as (5).

*Hydrostatic Equation:*

$$\frac{\partial M}{\partial \theta} = \Pi. \quad (6)$$

## 4. The ALE coordinate

FIM belongs to a category of circulation models referred to as *layer* models, meaning that vertical spacing of grid points is allowed to vary in space and time. Since the prognostic equations resemble the shallow-water equations – even in the sense that variables carried within layers are for many intents and purposes treated as vertically constant – layer models are also referred to as stacked shallow-water models.

The hybrid grid in FIM resembles that of RUC (Benjamin et al. 2004; Bleck and Benjamin 1993), but vertical staggering of variables is different because the RUC scheme does not rigorously conserve mass field constituents (referred to here as “tracers”). In RUC,  $\theta$  is carried on interfaces as in the Charney-Phillips grid (Arakawa and Moorthi 1988). This staggering convention, which is also followed in the Konor and Arakawa (1997) model, makes restoration of  $\theta$  to its prescribed coordinate value (see below) somewhat easier in RUC than in FIM which uses Lorenz type staggering (Arakawa and Moorthi 1988). FIM staggering replicates the layer treatment in the hybrid-isopycnic ocean model HYCOM (Bleck 2002) where only pressure and geopotential are carried on interfaces while all other variables, including tracers and Montgomery potential, are defined in layers.

The vertical coordinate in FIM is designed around the idea that coordinate layers conform to isentropic layers except in locations where these intersect the earth’s surface. There, layers are locally redefined as terrain-following ( $\sigma$  coordinate) layers. An individual coordinate layer can be isentropic in one geographic region and terrain-following in another.

The hybridization concept employed here and in RUC differs from hybrid schemes developed elsewhere (Bleck 1978; Konor and Arakawa 1997; Pierce et al. 1991; Webster et al. 1999; Zapotocny et al. 1991, 1994) in that it relies on locally mandated adjustment of vertical grid spacing rather than on a fixed formula typically consisting of a weighted average of two or more traditional coordinate choices. The vertical coordinate in FIM is best described as Arbitrary Lagrangian-Eulerian (Hirt et al. 1974), but we have added a mechanism for keeping coordinate layers aligned with their designated target isentropes over time when- and wherever this does not conflict with minimum layer thickness constraints. The original ALE scheme (*loc.cit.*) only concerned itself with the

maintenance of nonzero grid spacing in Lagrangian coordinate simulations. Coordinate “maintenance” in the sense of keeping coordinate surfaces aligned with iso-surfaces of some physical property is essential if an ALE circulation model is to be used in long-range weather or climate simulation.

While the flexibility of coordinate placement in ALE-type schemes is disconcerting to some users because grid point location in model space cannot be expressed in terms of a simple analytic formula, the ALE scheme excels in maximizing the size of the purely isentropic subdomain. This follows from the scheme’s ability to set the height where the  $\sigma$  coordinate gives way to the  $\theta$  coordinate in each geographic location separately, i.e., unencumbered by global considerations. Thus, while the lowest purely isentropic coordinate surface in schemes D of Bleck (1978) and Zapotocny et al. (1994) in a global model must be selected with the highest summertime temperature over the Tibetan Plateau in mind, the ALE approach contains no such restrictions. Furthermore, the problem of coordinate surface intertwining due to inappropriately chosen coordinate blending coefficients (Zhu et al. 1992) does not arise in an ALE scheme.

FIM manages the vertical grid structure as follows. If a given layer is “on target” (meaning that  $\theta$  matches the target potential temperature assigned to this layer) and if, in addition, the 2-D shallow-water continuity equation (eq. (4) without the  $\dot{s}$  terms)

$$\frac{\partial \Delta p}{\partial t} + \nabla \cdot (\mathbf{v} \Delta p) = 0 \quad (7)$$

yields a layer thickness  $\Delta p$  that does not fall below a predetermined minimum  $\Delta p_{min}$ , the  $\Delta p$  obtained from (7) is accepted. In other words, FIM sets  $\dot{s} = 0$  in this case, meaning that it treats the interfaces above and below the layer in question as material. If one of the above conditions is not met, the “grid generator” (see following section) takes over and changes  $\partial \Delta p / \partial t$  in a way that either maintains minimum thickness or, if the layer has become separated from its target potential temperature, moves it closer to it. In these situations, the selected  $\Delta p$  tendency is inserted into the full continuity equation (4) which at this point becomes a diagnostic equation for the interlayer mass fluxes  $\dot{s} \partial p / \partial s$ . The latter are used to vertically advect momentum and other variables.

The above process is complicated by the need to include conditions in several layers when deciding on the value of  $\dot{s} \partial p / \partial s$  on a given interface. However, as long



as  $\Delta p_{min} > 0$  is imposed only on layers at the bottom or top of the column, the test for nonzero  $\dot{s}\partial p/\partial s$  can be carried out recursively by a single sweep up or down the column, i.e., does not require iteration.

Hydrostatic models infer the vertical component of motion from the vertically integrated horizontal mass flux divergence. The grid generator in an ALE model divides this material vertical motion into two components: vertical motion of the coordinate surface and vertical air motion relative to it:

$$\begin{pmatrix} \text{vertical} \\ \text{motion} \\ \text{of} \\ \text{interface} \end{pmatrix} + \begin{pmatrix} \text{vertical} \\ \text{motion} \\ \text{through} \\ \text{interface} \end{pmatrix} = \begin{pmatrix} \text{vertically} \\ \text{integrated} \\ \text{horizontal} \\ \text{mass-flux} \\ \text{divergence} \end{pmatrix}. \quad (8)$$

The decision whether to accept the solution of (7) – that is, whether to balance the right-hand side of (8) by only the first term or by some combination of both terms on the left – is made by the grid generator at each grid point and each time step individually. The grid generator in FIM actually performs two tasks: it carries out the “regridding” just described, followed by a vertical “remapping” of all prognostic variables to the modified grid. The remapping is formally equivalent to vertical advection because it is driven by nonzero values of  $\dot{s}$ . However, since vertical displacement of atmospheric constituents due to actual air motion is already accounted for in the heaving and slumping of coordinate layers, the evaluation of the  $\dot{s}$  terms in the prognostic equations is best viewed as a secondary property redistribution initiated by the migration of coordinate surfaces through resting air. With the atmosphere conceptually “frozen” in time during this redistribution, remapping should conserve certain integral properties such as column integrals of momentum, thermal energy, etc.

Turbulent vertical mixing is typically parameterized in large-scale atmospheric models by solving a vertical diffusion equation with an eddy diffusivity coefficient designed to incorporate the effects of stratification, wind shear, etc. In an isentropic framework, one of the diffused variables is an independent variable, namely, the vertical coordinate  $\theta$ . The approach taken in FIM to deal with this peculiarity of isentropic modeling is described in the Appendix.

## 5. The Vertical Grid Generator

### 5a. Background

Owing to differences in the vertical staggering of variables, not all schemes existing today for maintaining the vertical grid structure in ALE-type hybrid-isentropic layer models are interchangeable. The scheme developed for RUC (Bleck and Benjamin 1993; Benjamin et al. 2004) in particular cannot be used directly in FIM. Instead, the technology suitable for the staggered FIM grid had to be imported from the ocean model HYCOM.

The first-generation HYCOM grid generator, whose design principles are described in detail in Bleck (2002) but date back to Bleck and Boudra (1981), has been modified and tuned over the years to address grid degeneracies that came to light as the range of applications of HYCOM grew. This tuning has added branches to the decision tree in the original algorithm, creating a situation where the underlying logic is no longer transparent to the user. Complexity in the grid generator discourages experimentation and adaptation of HYCOM/FIM to special modeling needs, and hence should be avoided.

The algorithm described below represents an attempt to get “back to basics” when moving layer interfaces for the sake of maximizing the part of the atmosphere represented by isentropic layers while at the same time subjecting the layers to minimum-thickness constraints. In the first-generation grid generator, each grid point is inspected and adjusted recursively in light of its distance to grid points above and below, using a variety of semi-empirical criteria. The algorithm proposed here is more straightforward in that it begins by transforming a given hybrid stairstep  $\theta$  profile into a purely isentropic one, i.e., into a stairstep profile whose  $\theta$  levels are prescribed beforehand. Depending on the stratification and  $\theta$  range in the original profile, this process can produce massless (zero thickness) layers at the top and bottom of the column. Massless layers that occur at the ground are subsequently inflated to a prescribed minimum thickness.

Differences between the resulting hybridized layer interface pressures and those of the input profile imply mass exchange among layers. Tracers and momentum must then be exchanged between layers as well. Any one of the standard conservative advection schemes can be used for this task.

### 5b. Transformation of non-isentropic stairstep $\theta$ profiles to isentropic coordinates

The following is an adaptation of the ocean-oriented scheme described in Appendix D of Bleck (2002).

Let  $\mathcal{Z}$  be a monotonic function of  $p$  decreasing with height, and let  $\theta_{in}(\mathcal{Z})$  be a piecewise constant (“stairstep”) vertical profile of  $\theta$ . Both the step width  $\Delta\theta_{in}$  and the step height  $\Delta\mathcal{Z}$  can vary from step to step. Our task is to transform  $\theta_{in}(\mathcal{Z})$  into another stairstep profile differing from the original one in that the location of the “risers” on the  $\theta$  axis is prescribed. Ideally, the transformation should be accomplished without perturbing the potential/internal energy of the column. Another quantity worth preserving is the geopotential height of the column, because a transformation that changes the column height is likely to set off external gravity waves.

Let  $\theta_k$  ( $k = 1, \dots, n$ ,  $\theta_{k+1} > \theta_k$ ,  $k$  increasing upward) mark the points on the  $\theta$  axis where we want the new risers to be placed. We require that the  $\theta_k$  values span the  $\theta$  range of the input profile,

$$\theta_1 \leq \theta_{in}(\mathcal{Z}) \leq \theta_n \text{ for all } \mathcal{Z}, \quad (9)$$

and that the input profile be monotonic. Denoting the pressure<sup>1</sup> of the lower and upper interface bounding the  $k$ -th layer by  $\mathcal{Z}_{k-1/2}$  and  $\mathcal{Z}_{k+1/2}$ , respectively, the condition we wish to satisfy can then be stated as

$$\sum_{k=1}^n \theta_k (\mathcal{Z}_{k-1/2} - \mathcal{Z}_{k+1/2}) = \int_{\mathcal{Z}_{n+1/2}}^{\mathcal{Z}_{1/2}} \theta_{in} d\mathcal{Z}. \quad (10)$$

The interface pressures are the unknowns in the problem.

Integration by parts (on the left this amounts to re-ordering the terms under the summation sign) allows us to rewrite (10) as

$$\begin{aligned} \theta_1 \mathcal{Z}_{1/2} - \theta_n \mathcal{Z}_{n+1/2} + \sum_{k=1}^{n-1} \mathcal{Z}_{k+1/2} (\theta_{k+1} - \theta_k) \\ = [\theta_{in} \mathcal{Z}]_{\mathcal{Z}_{n+1/2}}^{\mathcal{Z}_{1/2}} + \int_{\theta_{in}(\mathcal{Z}_{1/2})}^{\theta_{in}(\mathcal{Z}_{n+1/2})} \mathcal{Z}_{in}(\theta) d\theta \end{aligned}$$

where  $\mathcal{Z}_{in}(\theta)$  is the inverse of  $\theta_{in}(\mathcal{Z})$ .

In situations where the  $\theta$  range of the input profile does not span the entire range  $\theta_1 \dots \theta_n$ , we can, without altering the physical appearance of the input profile, lower

<sup>1</sup>While  $\mathcal{Z}$  is a function of pressure, we will refer to it as pressure for short.

$\theta_{in}(\mathcal{Z}_{1/2})$  to  $\theta_1$  and/or raise  $\theta_{in}(\mathcal{Z}_{n+1/2})$  to  $\theta_n$ . With these modifications, the above expression reduces to

$$\sum_{k=1}^{n-1} \mathcal{Z}_{k+1/2} (\theta_{k+1} - \theta_k) = \int_{\theta_1}^{\theta_n} \mathcal{Z}_{in}(\theta) d\theta. \quad (11)$$

Our strategy is to satisfy (11) by breaking the integral into pieces taken over intervals  $(\theta_k, \theta_{k+1})$  and conserving each integral individually. This immediately leads to

$$\mathcal{Z}_{k+1/2} = \frac{1}{\theta_{k+1} - \theta_k} \int_{\theta_k}^{\theta_{k+1}} \mathcal{Z}_{in}(\theta) d\theta \quad (12)$$

( $k = 1, \dots, n-1$ ).

If condition (9) is violated, evaluation of (12) is postponed until the offending input layer is brought into compliance by “diluting” it with mass from adjacent layers. Persistent heating at the model top, for example, is thereby transformed into a gradual thickening of the uppermost coordinate layer.

### 5c. Enforcement of layer thickness constraints

Suppose the prescribed potential temperature values  $\theta_1, \theta_2, \dots$  in the output profile cover a wide enough range to yield  $\theta_k < \theta_{in}$  for some  $k > 1$ . In this case, (12) will yield  $\mathcal{Z}_{1/2} = \mathcal{Z}_{3/2} = \dots = \mathcal{Z}_{k+1/2}$ , i.e., layers  $1, \dots, k$  in the transformed profile will be massless. Likewise, if  $\theta_{in} < \theta_k$  for some  $k < n$ , layers  $k+1, \dots, n$  will be rendered massless ( $\mathcal{Z}_{k+1/2} = \dots = \mathcal{Z}_{n+1/2}$ ).

The strategy in FIM is to accept massless layers aloft, but to always inflate massless layers at the bottom of the grid column. Layer inflation rules can be as simple as specifying a constant minimum thickness  $\Delta\mathcal{Z}_0$ . In this case the set of isentropic interface values  $\mathcal{Z}_{k+1/2}$  obtained from (12), to be identified here as  $\hat{\mathcal{Z}}_{k+1/2}$  to distinguish them from the final “hybridized” values, are recursively subjected to the constraint

$$\mathcal{Z}_{k+1/2} = \min(\hat{\mathcal{Z}}_{k+1/2}, \mathcal{Z}_{k-1/2} - \Delta\mathcal{Z}_0) \quad (13)$$

( $k = 1, 2, \dots$ ). Note that  $\Delta\mathcal{Z}_0$  can easily be made layer-dependent or scaled by terrain height.

It is advisable to smooth out large lateral variations in layer thickness that typically occur where a hybridized coordinate layer transitions from the fixed-depth to the isentropic subdomain. These variations are created when, for a given  $k$ , the  $2^{nd}$  argument in the *minimum*

function of (13) is chosen in one grid column, while the 1<sup>st</sup> argument is chosen in a neighboring column. One way to smooth out the transition, short of exchanging information among neighboring grid columns, is to increase layer thickness in situations where the two arguments are of similar magnitude. This is the purpose of the “cushion” function originally introduced into hybrid-coordinate ocean modeling by Bleck and Boudra (1981) and later adapted for atmospheric use by Bleck and Benjamin (1993). Use of the cushion function entails replacing (13) by

$$\mathcal{Z}_{k+1/2} = \min(\hat{\mathcal{Z}}_{k+1/2}, \mathcal{Z}_{k-1/2} - \text{cushn}[\hat{\mathcal{Z}}_{k-1/2} - \mathcal{Z}_{k+1/2}, \Delta\mathcal{Z}_0]). \quad (14)$$

In the two extreme cases where  $\hat{\mathcal{Z}}_{k-1/2} - \mathcal{Z}_{k+1/2}$  is either large negative or large positive compared to  $\Delta\mathcal{Z}_0$ , the cushion function is designed to replicate the functionality of (13). In other words,  $\text{cushn}(a, b)$  returns  $a$  if  $a \gg b$ , and it returns  $b$  if  $-a \gg b$  ( $b > 0$ ). In between the two extremes,  $\text{cushn}$  varies smoothly, returning values as high as  $2\max(a, b)$ . In many cases, this widens a layer if its potential temperature is close to target, thereby softening the lateral interface height contrast between locations where the underlying layer is isentropic and where it is not.

If more effective interface smoothing in the  $\sigma$ - $\theta$  transition region is deemed necessary, a sideways-looking smoothing algorithm may be required.

At the time of this writing, the minimum thickness value  $\Delta\mathcal{Z}_0$  is set as follows.

1. A default value  $\Delta\mathcal{Z}_0(k)$  is specified for each layer  $k$ . Typical values (stated here in pressure units for easier reference) are 3 hPa in the bottom layer, gradually increasing to 10 or 15 hPa in layers above.
2. In an attempt to mimic the vertical spacing of conventional  $\sigma$  coordinate layers, all  $\Delta\mathcal{Z}_0(k)$  are multiplied by the factor  $(p_{surf} - p_{top})/(1000 \text{ hPa} - p_{top})$  where  $p_{surf}$  is the surface pressure and  $p_{top}$  is the pressure level (400 hPa or smaller) where coordinate surfaces in a conventional  $\sigma$  coordinate model cease to be terrain-following.
3. Starting in the lowest layer and moving up the column,  $\mathcal{Z}_{k+1/2}$  is compared against the lesser of (14) and  $\mathcal{Z}_{1/2} - \sum_{n=1}^k \Delta\mathcal{Z}_0(n)$ . If it exceeds the minimum of these two values, it is replaced by that minimum.

This is done recursively, i.e., altered interface values affect the inflation test in layers above.

4. The lowest layer *not* in need of inflation is labeled  $k_{\sigma\theta}$ ; it marks the transition from the  $\sigma$  to the  $\theta$  sub-domain.
5. The upper interface of layer  $k_{\sigma\theta}$  stays fixed by definition, but very thin isentropic layers qualifying for inflation based on (14) can occur higher up in the atmosphere. To keep these from unnecessarily being inflated, the value  $\Delta\mathcal{Z}_0(k)$  is reduced in layers  $k_{\sigma\theta} + 1, \dots, k_{\sigma\theta} + 4$  by the factors 0.4, 0.2, 0.1, and 0.05, respectively. The factor 0.05 is also used in layers  $k > k_{\sigma\theta} + 4$ .

#### 5d. Vertical advection

The “regridding” process described in the preceding sections must be followed by a “remapping” process in which model variables are advected vertically in response to changes in interface pressure. Borrowing from HYCOM, vertical advection of momentum and tracers is currently handled by either the piecewise linear or the piecewise parabolic method (PLM, PPM) (van Leer 1974; Colella and Woodward 1984). All variables mentioned are remapped in  $p$  space to conserve their mass-weighted column integral.

Potential temperature is a special case. The regridding process described earlier yields a new  $\theta$  distribution that may be viewed as resulting from upstream or donor cell advection in  $\mathcal{Z}$  space. To suppress the numerical diffusivity implied by this low-order scheme, FIM actually discards the  $\theta$  field resulting from the regridding exercise and replaces it by a field advected by the same higher-order scheme that is used for the other prognostic variables.

There is a price to be paid for inferring the amount of mass transferred between layers from a piecewise constant  $\theta$  distribution, as is done in (10)–(12), and subsequently using a higher-order scheme to remap  $\theta$ . Neither will the slab of air arriving in a layer have the potential temperature needed to precisely restore that layer to target, nor will the transfer leave  $\theta$  in the donor layer unchanged. However, we find that repetitive use of the restoring algorithm allows layers to reach their target relatively quickly.

One advantage of using a higher-order advection

scheme for  $\theta$  is that the variable  $\mathcal{Z}$  in (10) – (14) no longer needs to be chosen with an eye on the conservation properties of the regridding scheme. Any variable monotonic in pressure, including  $p$  itself, is acceptable. What matters now is the vertical coordinate used during *remapping* of  $\theta$ . FIM allows the use of either  $(p/p_0)^\kappa$  or  $(p/p_0)^{1+\kappa}$  (where  $\kappa = R/c_p$ ). The rationale for providing these two options is given in the following section.

### 5e. Conservation alternatives

It follows from (6) that the height of an air column can be preserved during vertical regridding and remapping by setting  $\mathcal{Z} = \Pi$ . Unfortunately, this choice of  $\mathcal{Z}$  does not allow us to satisfy another important constraint: conservation of column-integrated internal energy  $I = \int c_v T dp$  and column-integrated potential energy  $P = \int g \rho z dz$ . (In an ideal gas, internal and potential energy are proportional to one another, so conservation of one entails conservation of the other.) The incompatibility of column height conservation with internal/potential energy conservation becomes clear if one writes  $P$  and  $I$  in terms of  $\theta$  and  $p$  and compares the resulting expressions

$$I = \frac{c_v p_0}{g(1+k)} \int \theta d \left( \frac{p}{p_0} \right)^{1+k} \quad (15)$$

$$P = \frac{R p_0}{g(1+k)} \int \theta d \left( \frac{p}{p_0} \right)^{1+k} \quad (16)$$

( $k = R/c_p$ ) with the formula for column height,  $\int \theta d\Pi$ , in which  $\theta$  is integrated over a variable proportional to  $(p/p_0)^\kappa$ . It is easy to see now that conservation of  $I$  and  $P$  can be achieved during remapping of  $\theta$  by using  $(p/p_0)^{1+k}$  as vertical coordinate, but that this can only be done at the price of violating the height preservation constraint. The relative importance of height versus internal/potential energy conservation is hard to assess without practical tests.

### 5f. Miscellaneous refinements

Discretization of a continuous profile  $\theta(p)$  in term of a staircase profile is not unique, because stairsteps can be broken into smaller steps or combined into bigger ones without violating any continuity or conservation principle. This ambiguity can lead to computational modes in the vertical layer structure, leading to the gradual disappearance of, say, odd-numbered layers accompanied

by a thickening of even-numbered ones. Initial experiments with FIM indeed revealed a propensity for amplifying this mode. To suppress it, a special algorithm has been added to the grid generator.

The algorithm scans each grid column for sequences of 5  $\Delta p$  values, numbered  $\Delta p_1, \dots, \Delta p_5$ , that satisfy the following three conditions:

$$\begin{aligned} \Delta p_1 &< \Delta p_2 \\ \Delta p_5 &< \Delta p_4 \\ \Delta p_3 &< \min(\Delta p_2, \Delta p_4). \end{aligned}$$

If all three conditions are met, layer 3 is inflated by drawing mass from both layers 2 and 4 such that (a) the column integral of  $\theta$  is conserved and (b)  $\Delta p_3^{new} = \min(\Delta p_2^{new}, \Delta p_4^{new})$ . Requirement (a) leads to the constraint

$$\frac{\Delta p_2 - \Delta p_2^{new}}{\Delta p_4 - \Delta p_4^{new}} = \frac{\theta_4 - \theta_3}{\theta_3 - \theta_2}$$

which may put a limit on the mass transfer stipulated by (b). The resulting interface displacements are added to those associated with the primary regridding process.

Suppression of the layer thickness computational mode improves the performance of the column physics parameterization scheme which has been found to be sensitive to large variations in layer thickness.

### 5g. An example

Some design options suitable for an ALE coordinate in atmospheric models are illustrated in Fig. 1. The figure shows a meridional cross section (latitude increasing to the left) that cuts across a typical midlatitude jet stream in thermal wind balance with sloping tropospheric isentropes. Three elements are combined in the figure. Solid lines running across each panel represent layer interfaces. Shaded contours represent zonal wind speed. Colors filling alternate spaces between isotachs show potential temperature. The purpose of the coloration is to indicate where in the domain a given coordinate layer is isentropic. The rendering is not exact because the columnwise steppy  $\theta$  field has been converted into a continuous field and interpolated to  $p$  space to simplify plotting.

The first panel in Fig. 1 shows a layer configuration typically seen in pure isentropic coordinate models like those of Bleck (1984) and Zhu et al. (1992). Since FIM evaluates lateral fluxes terms in (4) and (5) using the Flux

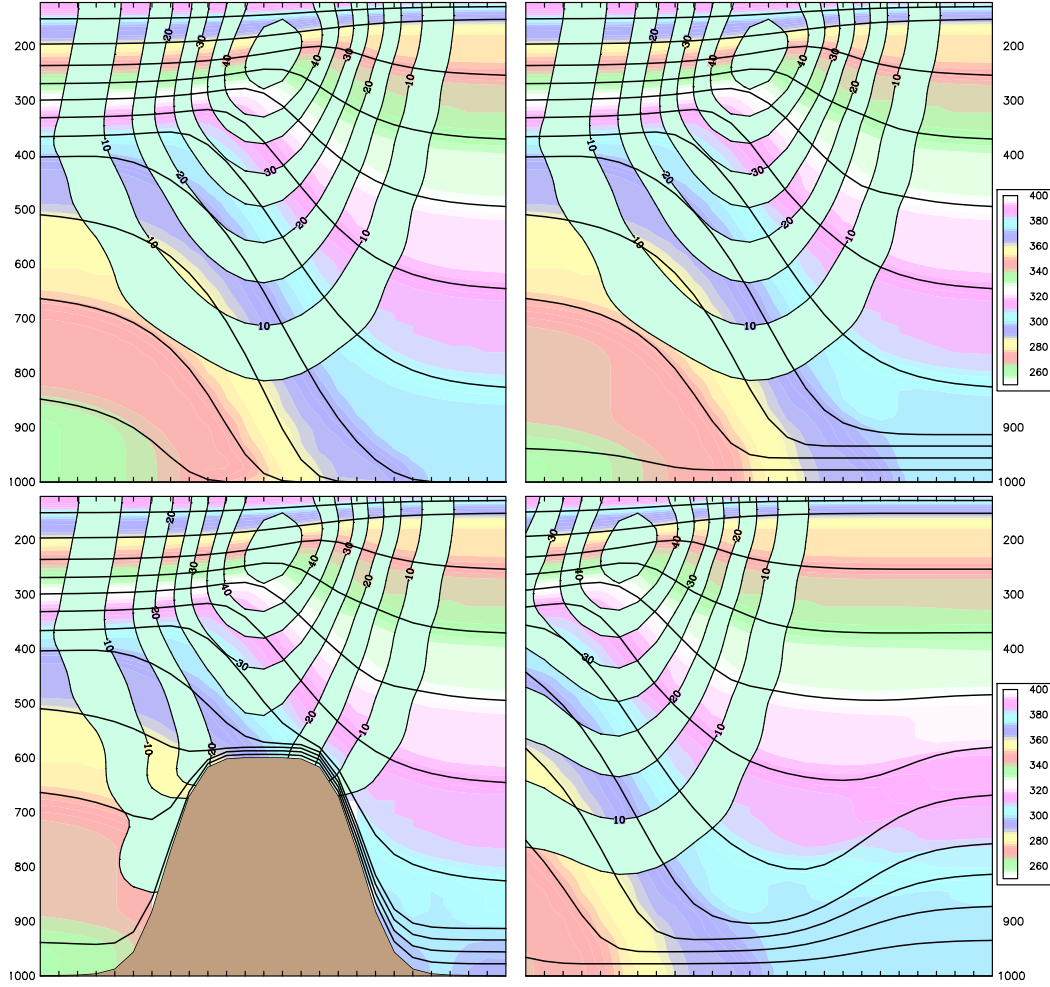


Figure 1: Vertical-meridional cross section illustrating the functionality of the ALE coordinate in FIM. Solid lines: layer interfaces Shaded contours: isotachs ( $\text{m s}^{-1}$ ). Color scheme: potential temperature (K). Ordinate: pressure (hPa). Horizontal extent: 5000 km

Corrected Transport scheme which permits layer thickness to go to zero (Lee et al. 2009), FIM can actually operate stably in the pure isentropic mode depicted in panel 1. The rationale for building this capability into FIM was to remove numerical-stability related constraints on the choice of minimum layer thickness.

Passing the configuration shown in panel 1 of Fig. 1 to the FIM grid generator yields the configuration shown in the second panel (upper right). As described in detail above, the grid generator inflates layers that intersect the ground – the ones shown as massless layers in the first panel – but leaves higher layers unmodified. It also tries to remove thickness variations from the lowest layer to aid in the calculation of surface fluxes. The shallow layers formed in this way near the earth’s surface are, of

course, no longer isentropic. A configuration similar to this one but with much higher vertical resolution is used in RUC (Benjamin et al. 2004) and in FIM. For illustrative purposes, minimum layer thickness  $\Delta p_{min}$  in this example is set to 30 hPa. FIM and RUC typically use values in the 10-20 hPa range.

The third panel (lower left) illustrates how the ALE coordinate reacts to the presence of a mountainous feature like the Tibetan Plateau. To keep the mountain from creating a nonisentropic coordinate cap extending to jet stream levels,  $\Delta p_{min}$  is reduced over high terrain as it would be in a  $\sigma$  coordinate model.

In the fourth panel we give an example of how one could modify the ALE coordinate to optimize layer spac-



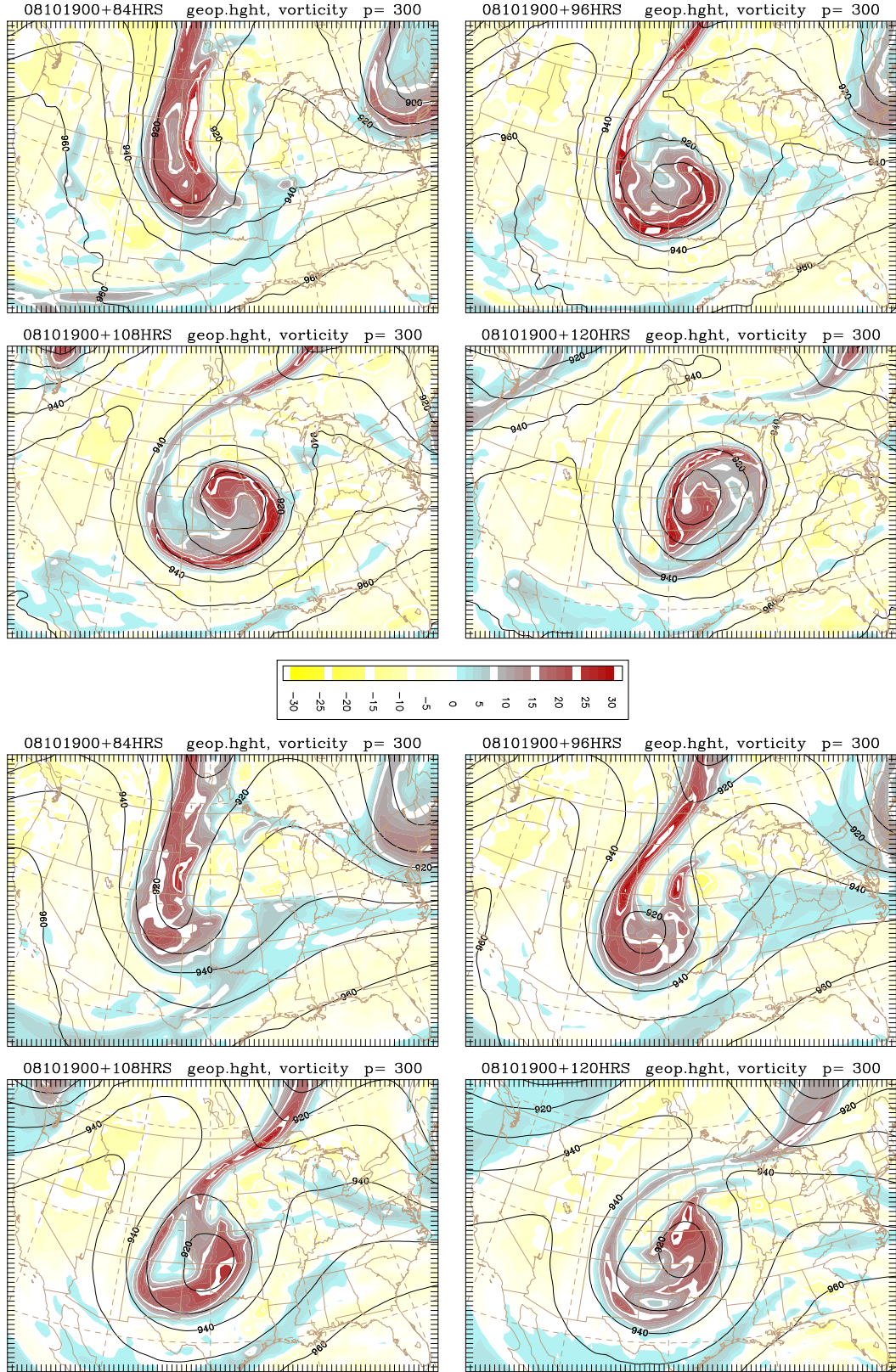


Figure 2: 84-, 96-, 108-, and 120-hr forecasts of geopotential height (m) and absolute vorticity ( $10^{-5}\text{s}^{-1}$ ) at 300 hPa. Initial time: 0000GMT, 19 Oct. 2008. Top 4 panels: forecasts based on FIM “native”  $\sigma-\theta$  grid. Bottom 4 panels: FIM forecasts using GFS  $\sigma-p$  grid. Horizontal resolution: G8 ( $\sim 30$  km).

ing at low latitudes where cloud physics parameterization schemes typically demand more uniform vertical grid spacing than that shown in panel 2. A layer expansion feature like the one in panel 4 would be particularly desirable in a model that uses an extremely low  $\Delta p_{min}$  for the sake of maximizing isentropic grid representation in the extratropics. Given the rather large value of 20 hPa currently used in FIM, refinements of this type are not contemplated at this time.

Because of the continuous rendering of the steppy  $\theta$  field, the color fields in Fig. 1 do not accurately reflect  $\theta$  in each layer, as already mentioned. However, some color irregularities in regions where interfaces have been pushed up or down by the grid generator are the result of vertical advection errors.

This brings up an important point. The flexibility of the ALE scheme might compel a model architect to design a coordinate system that requires interfaces to periodically move over large distances through the fluid. Such a design, an obvious violation of the “flow-following” concept, can lead to vertical advection and associated dissipation errors far in excess of those typically encountered in a fixed grid. One example of an ill-advised ALE algorithm is to attach a coordinate surface to the top of the planetary boundary layer for the sake of cleanly separating well-mixed from stratified air layers. Due to the potentially large difference between daytime and nighttime boundary layer height and the ensuing need to move large amounts of air back and forth across the pulsating interface, advantages gained by separating regions of large and small vertical mixing are likely to be lost. The best strategy is to make coordinate surfaces either truly Lagrangian or, if that is impractical, keep them fixed in space.

There are situations where not only time- but also space-dependent  $\Delta p_{min}$  values of the kind shown in panel 4 of Fig. 1 must be avoided. Consider, for example, a model in which the Sadourny (1975) or Arakawa and Lamb (1981) approach is used to formulate the momentum equations in potential-entropy conserving form, with layer thickness  $\Delta p$  taking the place of  $\partial p / \partial \theta$  in the potential vorticity expression  $(\zeta + f)(\partial p / \partial \theta)^{-1}$ . In the  $\sigma$  coordinate subdomain, lateral variations imposed on  $\Delta p$  by the grid generator will in this case introduce bogus lateral potential vorticity gradients which can distort the flow evolution inappropriately. FIM avoids this pitfall by using vorticity, not potential vorticity, in (3). HYCOM

uses the Sadourny (1975) formulation but makes  $\Delta p_{min}$  a function of layer index only, effectively replacing potential vorticity by absolute vorticity inside the  $p$  coordinate subdomain.

## 6. A sensitivity study

A displaced fluid parcel that does not experience a buoyancy force driving it back to its original location is likely to remain in contact with its new surroundings longer than it otherwise would be. This gives the parcel some extra time to exchange properties with the surrounding fluid. Consequently, turbulent mixing in stratified fluids takes place preferentially along surfaces of constant potential buoyancy (i.e. buoyancy corrected for compressibility effects).

With potential buoyancy and entropy being synonymous (or nearly so) in the atmosphere, turbulent exchange tends to minimize isentropic gradients of properties such as momentum (Rossby and Collaborators 1937). Replicating this process in a numerical model is not easy if coordinate surfaces do not coincide with isentropes. Dispersion errors associated with horizontal transport in a model tend, over time, to destroy property contrasts on whichever surfaces the transport is being carried out. Only if the transport equations are solved on isentropic surfaces can this numerical dispersion error be hidden behind what we may call a “smokescreen” of naturally occurring mixing. Being able to do this is especially advantageous in long-term simulations of statistically stationary or slowly varying states that in nature result from a balance between external forcing, transport, and turbulent mixing.

Note that the above argument refers to numerical errors associated with *lateral transport* – that is, errors caused by the dispersive properties of the horizontal advection operator. Errors arising during evaluation of explicit mixing terms (which FIM is free of) can also be important. However, these can be reduced – to some extent at least – by aligning the main axes of the mixing tensor with isentropic surfaces (Redi 1982). An equivalent strategy to project the effect of numerical dispersion in the transported field onto isentropic surfaces has not yet been developed.

The flexibility of the ALE coordinate allows us to shed light on the correctness of the assertion that numerical

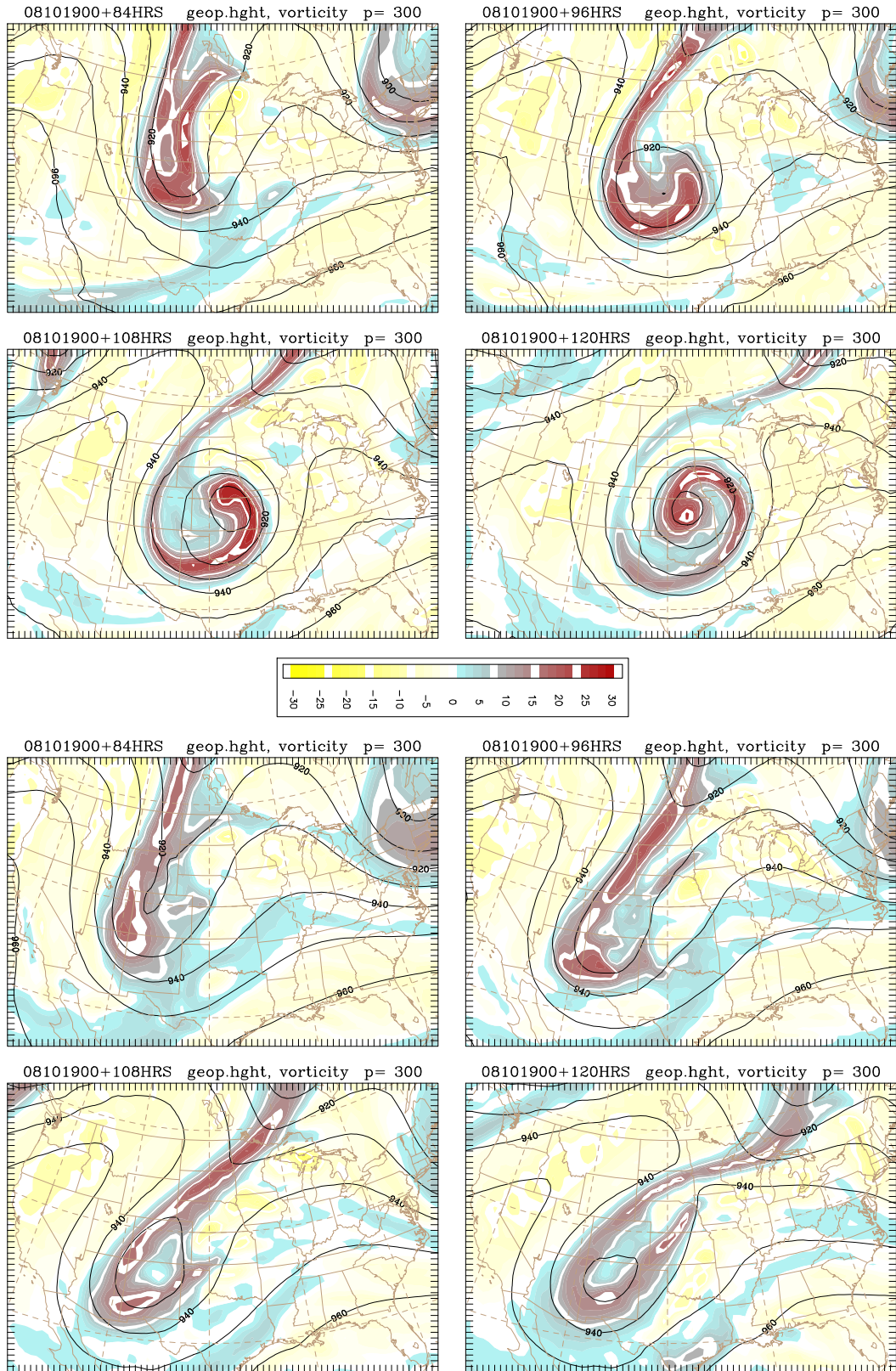


Figure 3: As in Fig. 2 but for G7 (~60 km) resolution.



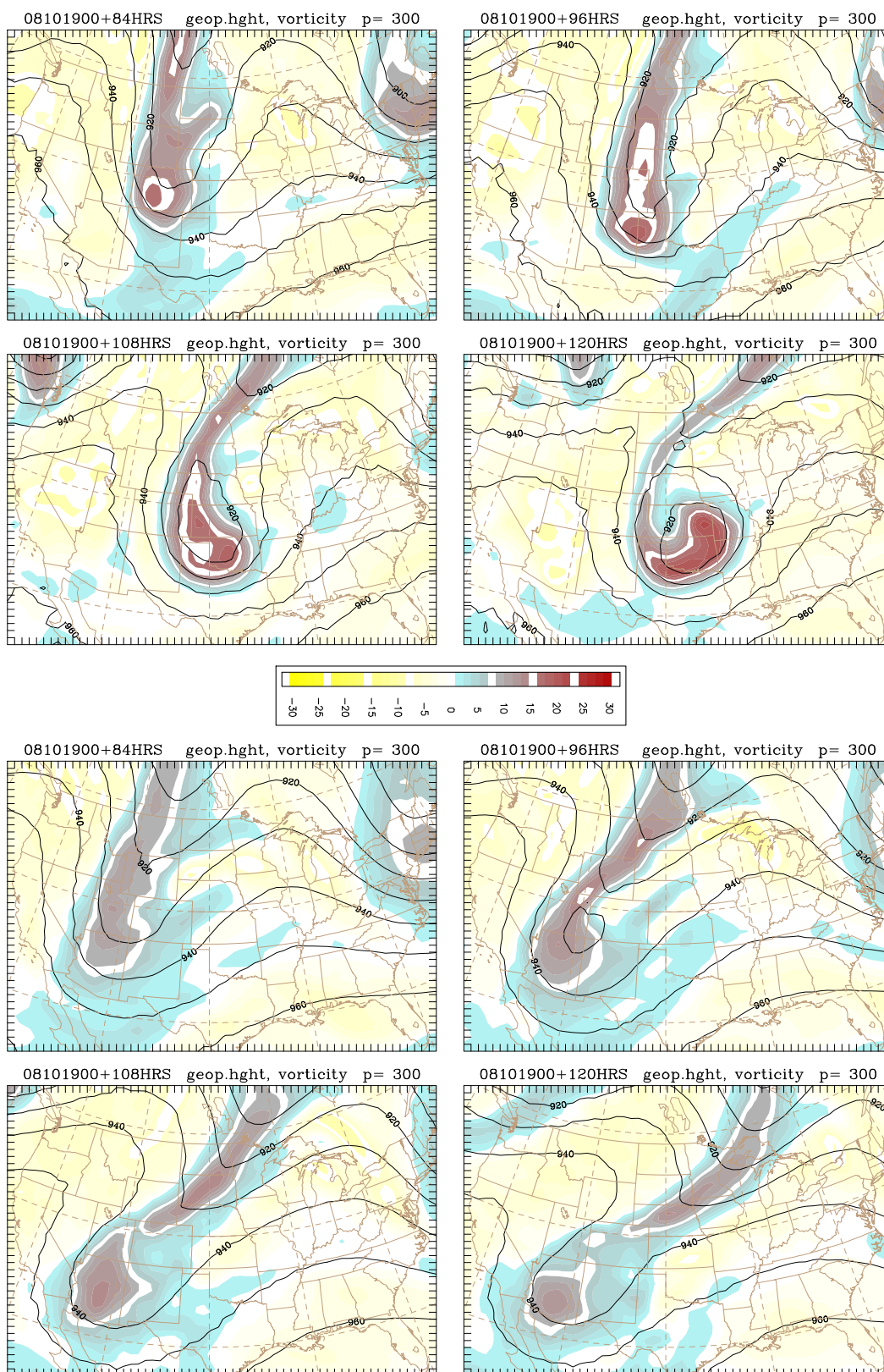


Figure 4: As in Fig. 2 but for G6 (~120 km) resolution.

accuracy of transport processes benefits from isentropic coordinate representation. We will do this by changing the vertical coordinate in FIM from hybrid  $\sigma$ - $\theta$  to a more traditional combination of  $\sigma$  and  $p$  and will look for at least anecdotal evidence that simulations based on the first-mentioned coordinate yield more coherent patterns of dynamically relevant quantities than simulations based on the latter.

A tracer well suited for this purpose is the absolute vorticity  $\eta = \partial v / \partial x - \partial u / \partial y + f$ . Even though  $\eta$  is neither explicitly advected in primitive equation models, nor is it rigorously conserved (especially if the  $u, v$  derivatives are taken at  $\sigma = \text{const}$  as opposed to  $\theta = \text{const}$ ), the fact that vorticity is composed of spatial derivatives of the velocity field and interacts with the circulation in a two-way mode makes it a particularly sensitive indicator of forecast errors. We will focus in the following on the process of tropospheric Rossby wave breaking (McIntyre and Palmer 1985), also referred to as vortex rollup (Dritschel and Polvani 1992) or, in synoptic meteorologists' parlance, cutoff low formation.

Three synoptic cases were analyzed in detail for this article. Space limitations permit us to present only one of them in detail. Limited results from the other cases will be included to indicate that trends in forecast accuracy as functions of horizontal resolution and vertical grid were not unique to the first case.

The vortex rollup process, and the failure of some model versions to simulate it accurately, were found to be depicted most succinctly by the vorticity and geopotential height distribution on the 300 hPa isobaric surface. We will therefore confine our attention to 300 hPa flow patterns and will first focus on forecasts over North America extending 3.5 to 5 days beyond the initial time of 0000GMT, 19 Oct. 2008. (The other two cases depict vortex rollup events over Europe four weeks later and over the Southern Ocean in late austral summer.)

A total of eight FIM forecasts were carried out, four using the native FIM coordinate and four using the GFS  $\sigma$ - $p$  coordinate. The switch from the former coordinate to the latter is accomplished in FIM by replacing the standard hybrid-isentropic grid generator by one that simply restores interface pressures to the values prescribed by the GFS. No other changes are made to the model.

The forecasts within each group of four differ by horizontal grid resolution. As outlined in Sec.3a of Lee and

MacDonald (2009), recursively bisecting the sides of the 20 triangles in the icosahedron quadruples the number of hexagonal cells on the sphere. In the experiments reported here, the number of cells ranges from  $\sim 10,000$  (referred to as G5 resolution – 5 bisecting steps) to  $\sim 655,000$  (G8, 8 bisecting steps). The mesh size is  $\sim 240$  km at G5 and  $\sim 30$  km at G8 resolution; mesh size varies by approximately 15% over the sphere.

The figures for case 1 are organized as follows. Results obtained by FIM configured with its standard hybrid  $\sigma$ - $\theta$  coordinate are displayed in the top 4 panels of Figs. 2-5. Corresponding results obtained by substituting the hybrid  $\sigma$ - $p$  coordinate for the “native” FIM coordinate are shown in the bottom 4 panels of each figure. Coordinate values in the  $\sigma$ - $p$  grid are identical to those used by the Global Forecast System at the NOAA National Centers for Environmental Prediction. Both grids used here consist of 64 layers.

We begin by showing in Fig. 2 the rollup process as simulated at the highest available resolution of 30 km (G8). The forecasts clearly depict the process by which mixing in fluids takes place: initially compact fluid elements are continually deformed into long, thin filaments that create sharp property gradients subsequently eroded by molecular diffusion. How important explicit simulation of this stretching or stirring process is for climate modeling is an important open question. There is no doubt, however, that such explicit simulation is important in numerical weather prediction because property gradients formed by filamentation often spawn severe local weather events.

Fig. 2 indicates that the  $\sigma$ - $\theta$  forecast carries the filamentation process somewhat farther in time than the  $\sigma$ - $p$  forecast. In fact, it may carry it too far. Comparison to observed conditions on 23 Oct. 2008 (not shown here) indicates that FIM using its native coordinate actually over-intensifies the cutoff vortex. Further experimentation will be required to determine whether omission of explicit subgridscale mixing terms in FIM is an occasional detriment to forecast accuracy.

The following figures depict the rate at which the filamentation and rollup process is degraded with lower grid resolution. At G7 (60 km) resolution (Fig. 3), the vorticity streamer in the  $\sigma$ - $\theta$  forecast is seen to maintain its integrity, even to the extent that its spiral structure at 120 hrs is better defined in this forecast than in the previous one. Larger changes between Fig. 2 and Fig. 3

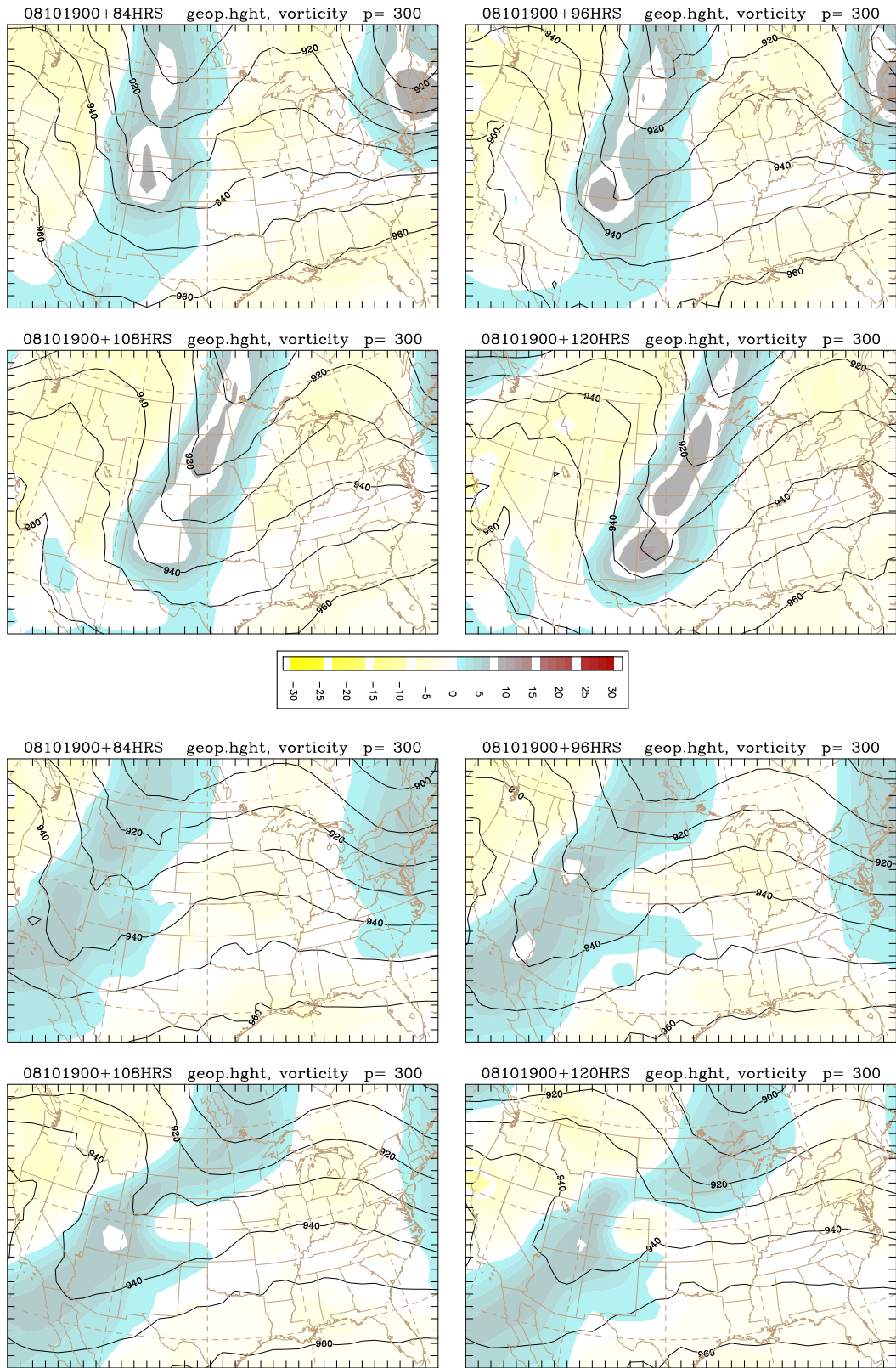


Figure 5: As in Fig. 2 but for G5 (~240 km) resolution.



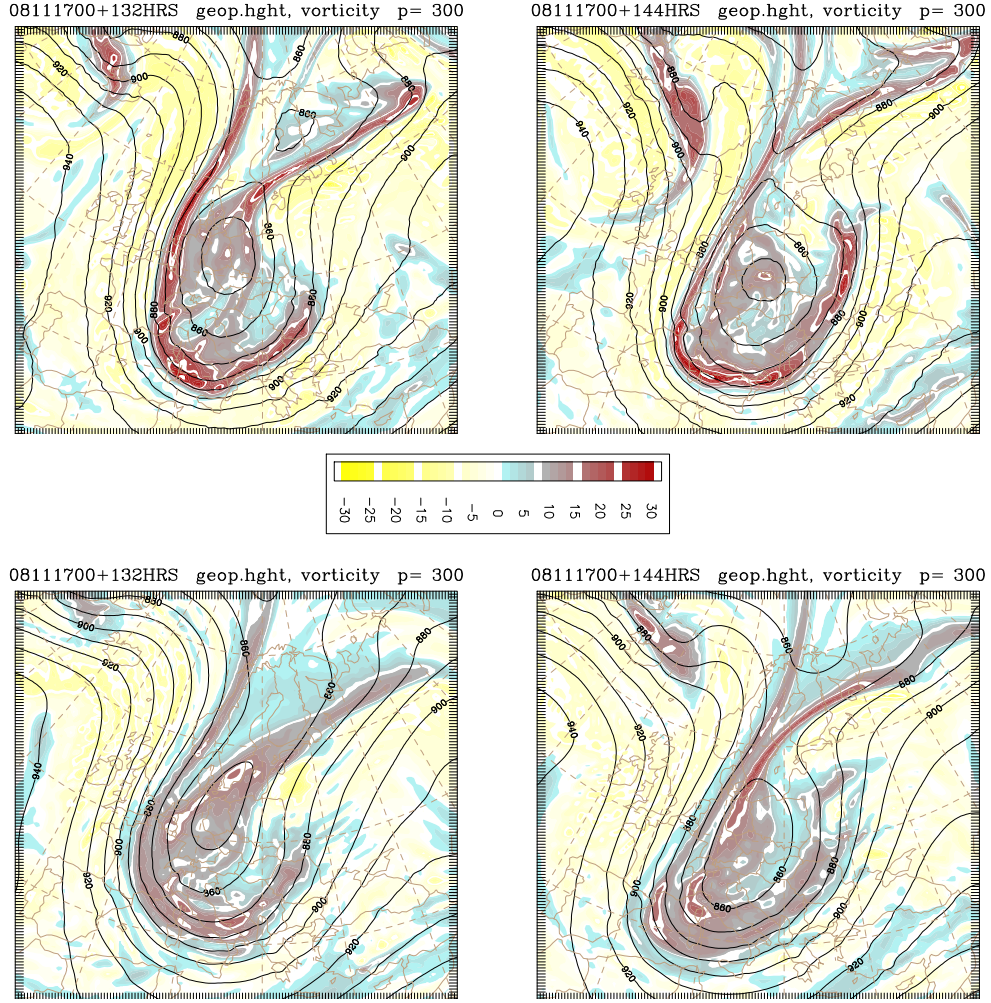


Figure 6: 132-, and 144-hr forecasts of geopotential height (m) and absolute vorticity ( $10^{-5}\text{s}^{-1}$ ) at 300 hPa. Initial time: 0000GMT, 17 Nov. 2008. Top: forecasts based on FIM “native”  $\sigma-\theta$  grid . Bottom: FIM forecasts using GFS  $\sigma-p$  grid. Horizontal resolution: G8 ( $\sim 30$  km).

are noticeable in the  $\sigma-p$  forecast, and these changes foreshadow a rather precipitous decline in filament definition with decreasing grid resolution in  $\sigma-p$  mode. The 300 hPa surface at 120 hrs in the  $\sigma-p$  forecast has risen by 10 m as a result of the resolution change, and the vortex is located too far to the west.

It is worth noting that even in a perfect model simulation the filament patterns in  $\sigma-\theta$  and  $\sigma-p$  forecasts would differ because the plotted vorticity is based on winds differentiated at constant  $\theta$  in one model version and at constant  $p$  in the other. (At 300 hPa, the GFS coordinate is nearly isobaric.)

The trend suggested by Figs. 2 and 3 continues as mesh size is doubled again, to 120 km (Fig. 4). At this

resolution the streamer in the  $\sigma-\theta$  forecast loses its distinctive spiral character, but its forward edge still shows signs of being wrapped around the vortex at 120 hrs. In the  $\sigma-p$  forecast at 120 km resolution, the vorticity field is devoid of sharp maxima and the rollup process is greatly weakened. The trough line develops a serious tilt toward the southwest.

At 240 km spatial resolution (Fig. 5) the  $\sigma-\theta$  forecast finally shows signs of serious degradation, reminiscent of what we saw at higher resolution in the  $\sigma-p$  forecast. Vortex rollup is no longer taking place, even though the erroneous tilt of the trough line is less serious in Fig. 5 (top panels) than in the 120 km  $\sigma-p$  forecast (bottom panels of Fig. 4). The trough in the 240 km  $\sigma-p$  forecast





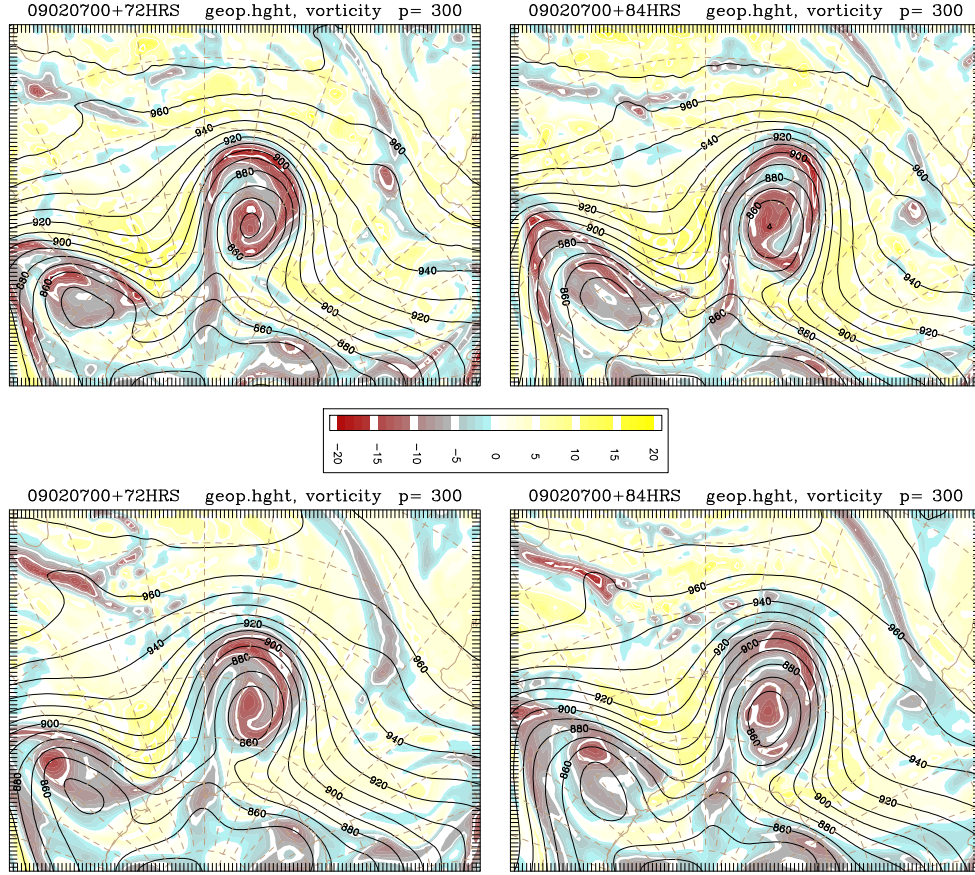


Figure 8: 72-, and 84-hr forecasts of geopotential height (m) and absolute vorticity ( $10^{-5}\text{s}^{-1}$ ) at 300 hPa. Initial time: 0000GMT, 7 Feb. 2009. Top: forecasts based on FIM “native”  $\sigma-\theta$  grid . Bottom: FIM forecasts using GFS  $\sigma-p$  grid. Horizontal resolution: G8 ( $\sim 30\text{km}$ ). Maps are cropped to show pieces of Madagascar and Australia for orientation.

Differences in the large-scale vorticity pattern between  $\sigma-p$  and  $\sigma-\theta$  forecasts are minor at this high resolution, even though the vorticity filaments appear to be slightly more coherent in the  $\sigma-\theta$  forecast.

The weakening of the wave-breaking process with decreasing horizontal resolution is not as pronounced as in the previous cases. Despite the relatively low resolution of 120 km in the G6 forecasts shown in Fig. 8, the remnants of a vorticity spiral are still very noticeable at 72 hrs in the  $\sigma-\theta$  forecast (upper left panel) whereas there is nothing left of it in the  $\sigma-p$  forecast. As in the November 2008 case, the “amount” of vorticity trapped in the vortex appears to be higher in the  $\sigma-\theta$  forecast, leading to a slightly stronger cutoff vortex.

An alternate view of the difference between the two forecasts is provided in Fig. 10. The cross sections shown correspond in time and grid resolution to the fields

in Fig. 9. The sections slice through the vortex in zonal direction, thereby creating a butterfly pattern in the velocity field. The jet in the  $\sigma-\theta$  forecast is the stronger one, roughly  $5\text{ms}^{-1}$  on the upstream side and  $10\text{ms}^{-1}$  on the downstream side. The difference in strength is conceptually consistent with the fact that the doming of isentropes in the vortex center reduces velocity gradients along coordinate surfaces on the cyclonic side of the jet, thereby lessening numerical dissipation of cyclonic vorticity below jet stream level.

## 7. Discussion

Evidence we have accumulated so far, though still largely anecdotal, suggests that FIM using its native  $\sigma-\theta$  coordinate maintains the integrity of tropospheric vorticity filaments better with decreasing resolution than the same

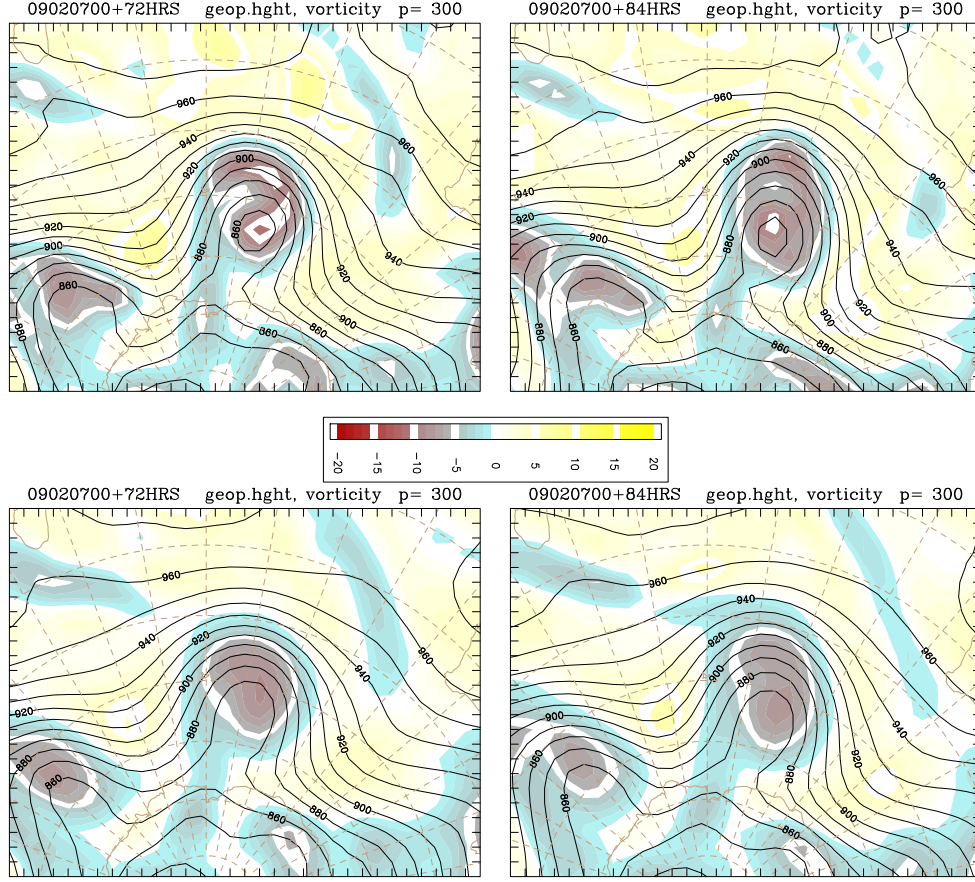


Figure 9: As in Fig.8 but for G6 resolution ( $\sim 120$  km).

model using a  $\sigma-p$  coordinate adopted from NCEP's Global Forecast System. The evidence is based on three extensively analyzed cases and on a few more cases analyzed in less detail, all chosen from the time period fall 2008 through winter 2009.

Nevertheless, the reasoning for why different trends in the simulation of Rossby wave breaking in the two model versions are to be expected is straightforward. As argued earlier, *numerically* induced lateral mixing of fine structures generated by filament stretching is least detrimental to forecast accuracy in models whose mixing surfaces coincide with iso-surfaces of potential buoyancy. FIM with its  $\sigma-\theta$  grid represents such a model, at least above the lowest few kilometers of the atmosphere.

One seeming flaw of the above case study is the juxtaposition of two different vorticity fields – one,  $\eta_\theta$ , based on differentiation at constant  $\theta$  and one,  $\eta_p$ , on differentiation at constant  $p$ . For this reason, comparison of  $\eta_p$  with  $\eta_\theta$  at the same spatial resolution is less meaningful

than a comparison of the resolution dependence of either the  $\eta_\theta$  or the  $\eta_p$  field by itself. We should therefore refrain from making conclusions such as “filament simulation in the  $\sigma-\theta$  model at  $G_n$  resolution (where  $n$  is a natural number) compares accuracy-wise to simulation in the  $\sigma-p$  model at resolution  $G_{n+1}$ .” On the other hand, it seems permissible to conclude that the deterioration of forecast accuracy with decreasing resolution is more rapid in the  $\sigma-p$  simulations than the  $\sigma-\theta$  simulations.

Even if quantitative information cannot be drawn from it, the exercise of plotting  $\eta_\theta$  on  $p$  surfaces and comparing the resulting distribution with  $\eta_p$  plotted on the same surfaces is still informative. It illustrates the level of detail in a dynamically relevant tracer field that in an isentropic coordinate model, by virtue of coordinate surface orientation, is largely shielded from nonphysical dissipation. The  $\eta_p$  field in the FIM version featuring a  $\sigma-p$  coordinate, on the other hand, directly feels the impact of numerical dissipation and deteriorates with time accordingly. It remains to be seen, of course, whether some

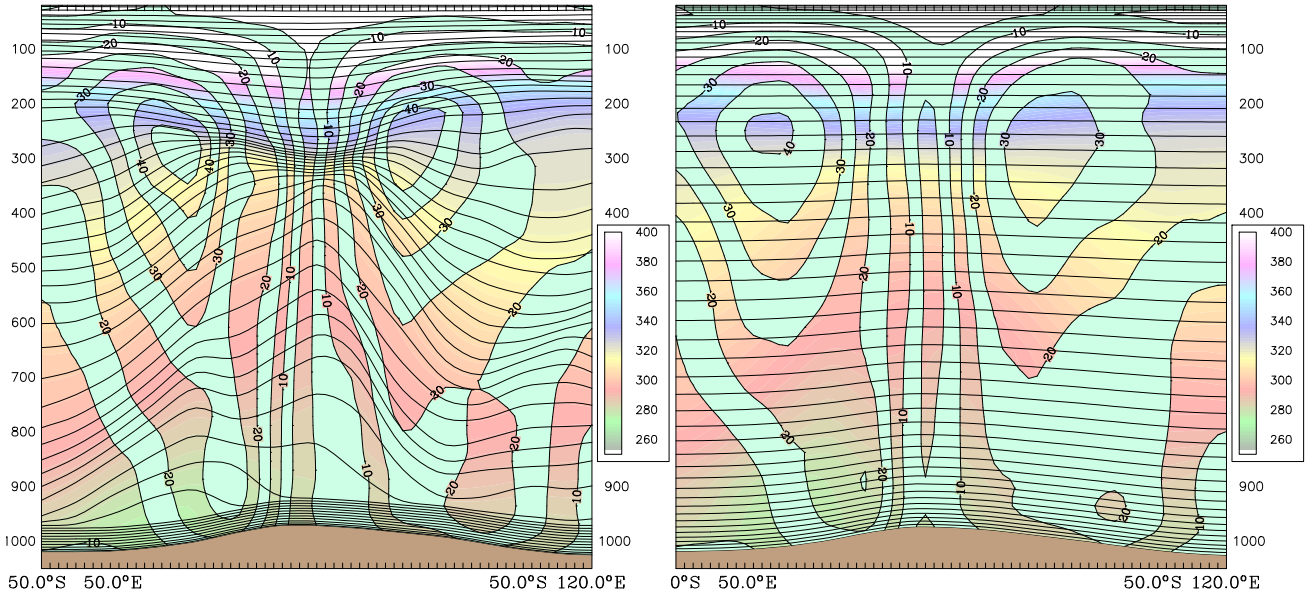


Figure 10: Zonal cross sections at 50S, 50E – 120E, extracted from 72-hr G6 forecasts based on  $\sigma-\theta$  coordinate with 5 hPa minimum layer thickness (left) and GFS  $\sigma-p$  coordinate (right). Features shown in the sections are explained in Fig. 1.

degree of deterioration might not be realistic. In other words, the question will have to be addressed whether the FIM equations should retain their present frictionless form or be amended by explicit mixing terms.

## 8. Closing remarks

Material has been presented supporting the notion that “flow-following” or quasi-Lagrangian vertical coordinates are a viable alternative to the Eulerian coordinates commonly used in atmospheric circulation modeling. This point, of course, has been made before by members of the Wisconsin and UCLA schools [Johnson (2000), Arakawa (2000), Randall et al. (2000), among others]. What is new about the present effort is that for the first time a quasi-isentropic coordinate model is being used on a routine basis for real-time, medium-range global weather prediction. Comparing FIM with the other two hybrid-isentropic models routinely used for weather prediction today, FIM differs from RUC (Benjamin et al. 2004) primarily in the use of an icosahedral global grid and from the University of Wisconsin global model (Schaak et al. 2004) in the use of an ALE coordinate.

Not all components of FIM are new. Model “physics”

routines were made available to FIM developers by the group responsible for the Global Forecast System at the NOAA National Centers for Environmental Prediction; FIM development efforts have greatly benefitted from this collaboration. Likewise, the global fields used to initialize FIM are imported directly from NCEP.

This said, FIM is unique in having combined two novel approaches to numerical weather prediction: (1) icosahedron- and finite volume-based horizontal discretization and (2) an entropy-based vertical coordinate. The present article documents the latter aspect in detail. In addition, material is presented suggesting that numerical diffusion attributed to the dispersive effects of the horizontal transport operators can be rendered less destructive by solving the dynamic equations in an isentropic coordinate system. The question of whether this by itself leads to improved forecast accuracy has not been explored in this article but will be the subject of future work.



## 9. Appendix: Turbulent Vertical Mixing

The following is a simplified version of a numerical scheme developed by McDougall and Dewar (1998) for carrying out vertical mixing in fluid models whose vertical coordinate is a function of the diffused variable(s). They deal with the specific problem of mixing temperature and salinity in ocean models whose vertical coordinate is potential density (a function of both temperature and salinity), constrained to remain constant in each coordinate layer during mixing.

Here we deal with the much simpler problem of solving the diffusion equation in an atmospheric column where there is only one diffused variable (potential temperature  $\theta$ ) doing double duty as vertical coordinate. The only variable capable of capturing the effects of thermal diffusion in this case is the thickness of coordinate layers.

The equations expressing conservation of mass and heat in a column, basically 1-D versions of the equations listed in the beginning, are

$$\frac{\partial}{\partial t} \left( \frac{\partial z}{\partial s} \right)_s + \frac{\partial}{\partial s} \left( \dot{s} \frac{\partial z}{\partial s} \right) = 0. \quad (17)$$

$$\left( \frac{\partial \theta}{\partial t} \right)_s + \left( \dot{s} \frac{\partial z}{\partial s} \right) \frac{\partial \theta}{\partial z} = - \frac{\partial F_\theta}{\partial z} \quad (18)$$

The turbulent heat flux  $F_\theta = \overline{w'\theta'}$  is usually parameterized as  $F_\theta = -K\partial\theta/\partial z$  where  $\theta$  is the resolved-scale potential temperature and  $K$  is a thermal diffusivity coefficient.

The flux form of (18), obtained by combining (17) and (18), is

$$\frac{\partial}{\partial t} \left( \theta \frac{\partial z}{\partial s} \right) + \frac{\partial}{\partial s} \left( \dot{s} \frac{\partial z}{\partial s} \theta \right) = - \frac{\partial F_\theta}{\partial s}. \quad (19)$$

The task at hand is to discretize the above equations for use in a model framework where stratification is represented by a piecewise constant, staircase  $\theta$  profile. The discretization will be done by formally integrating the equations over individual stairsteps.

If  $\theta$  is to remain constant in each layer during the mixing process,  $F_\theta$  must be vertically constant in each layer. If this were not the case, integrating (18) over an individual layer would yield a nonzero right-hand side. Of the two terms on the left, the second one integrates to zero

since  $\partial\theta/\partial z = 0$  inside the layer. (The vertical mass flux ( $\dot{s}\partial z/\partial s$ ) remains finite.) Hence, a nonzero r.h.s. implies a nonzero tendency term  $\partial\theta/\partial t$  which clashes with the stated requirement.

We conclude: for diffusion to leave a mark on the profile under the constraint  $F_\theta = \text{const}$  in individual layers,  $F_\theta$  must be allowed to vary from layer to layer. The implied infinite heat flux divergence at layer interfaces is consistent with the notion that air crossing an interface undergoes an instantaneous change in  $\theta$ .

A simple centered finite-difference expression for the heat flux in layer  $n$  is

$$F_\theta^n = \frac{K^n}{2} \frac{\theta^{n+1} - \theta^{n-1}}{z^{n+1/2} - z^{n-1/2}} \quad (20)$$

where fractional superscripts indicate quantities defined on interfaces.

The central task is to determine the mass flux across layer interfaces, ( $\dot{s}\partial z/\partial s$ ). For this we integrate (19) over an  $s$  interval representing an infinitesimal slab bracketing a layer interface. Since the tendency term drops out as  $\partial z$  approaches zero and the mass flux ( $\dot{s}\partial z/\partial s$ ) is continuous in the vertical, we obtain in the limit of zero slab thickness

$$\left( \dot{s} \frac{\partial z}{\partial s} \right)^{n+1/2} = \frac{F_\theta^{n+1} - F_\theta^n}{\theta^{n+1} - \theta^n}. \quad (21)$$

Expressions (20) and (21) encompass the sought-after solution to the problem of diffusing heat in a staircase  $\theta$  profile while maintaining  $\theta$  in individual layers. Note that, in the absence of externally imposed heat fluxes, the column integral  $\int \theta dz$  is conserved regardless of the physical and numerical approximations made in evaluating the heat flux (20).

The heat flux as approximated by (20) becomes infinite in massless layers. To avoid division by zero, the denominator in (20) must therefore be bounded away from zero. The parameter representing minimum layer thickness, together with  $K$  and the time step used in solving (19), can be tuned to concentrate the effect of vertical diffusion almost entirely on very thin layers. We use the scheme in this mode as an alternative to the grid generator to avoid generating zero-thickness layers in the isentropic subdomain that may result from strongly layer-dependent diabatic forcing. The advantage of the present scheme over the grid generator is that it does not produce local deviations from target  $\theta$ . Suitable parameter values are:  $1 \text{ m}^2$  for the product of time step and

mixing coefficient, and  $2 \times 10^{-3}\text{m}$  for the minimum thickness.

Ideally, vertical mixing should conserve the total heat content of the column,  $\int c_v T dp$ . From (15) we note that in order to conserve total heat, the variable  $z$  in (17) – (21) must be replaced by a variable proportional to  $p^{1+k}$ . No other changes are required in the solution procedure, except that the mixing coefficient  $K$  in (20) must be rendered dimensionally compatible with the new vertical coordinate.

In a similar vein, preservation of the total height  $\int \theta d\Pi$  of the column during mixing can be achieved by using a variable proportional to  $p^k$  in place of  $z$ . Note that height preservation is incompatible with heat conservation.

## References

- Arakawa, A., 2000: Future Development of General Circulation Models. In: General Circulation Model Development (D. A. Randall, ed.). *Internat. Geophys. Series*, **70**, Academic Press, New York, 721–780.
- Arakawa, A. and V. R. Lamb, 1981: A potential enstrophy- and energy-conserving scheme for the shallow water equations. *Mon. Wea. Rev.*, **109**, 18–36.
- Arakawa, A. and S. Moorthi, 1988: Baroclinic instability in vertically discrete systems. *J. Atmosph. Sci.*, **45**, 1688–1707.
- Benjamin, S., G. Grell, J. Brown, T. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the ruc hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494.
- Bleck, R., 1974: Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Wea. Rev.*, **102**, 813–829.
- 1978: On the use of hybrid vertical coordinates in numerical weather prediction models. *Mon. Wea. Rev.*, **106**, 1233–1244.
- 1984: An isentropic coordinate model suitable for lee cyclogenesis simulation. *Riv. Meteorol. Aeronaut.*, **44**, 189–194.
- 2002: An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, **4**, 55–88.
- Bleck, R. and S. Benjamin, 1993: Regional weather prediction with a model combining terrain-following and isentropic coordinates. part 1: model description. *Mon. Wea. Rev.*, **121**, 1770–1785.
- Bleck, R. and D. B. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.*, **11**, 755–770.
- 1986: Wind-driven spin-up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Geophys. Res.*, **91C**, 7611–7621.
- Boris, J. P. and D. L. Book, 1973: Flux-corrected transport. i. shasta, a fluid transport algorithm that works. *J. Comput. Phys.*, **11**, 38–69.
- Colella, P. and P. Woodward, 1984: The piecewise parabolic method (ppm) for gas-dynamical simulations. *J. Comput. Phys.*, **54**, 174–201.
- Deaven, D. G., 1976: A solution for boundary problems in isentropic coordinate models. *J. Atmosph. Sci.*, **33**, 1702–1713.
- Dritschel, D. G. and L. M. Polvani, 1992: The roll-up of vorticity strips on the surface of a sphere. *J. Fluid Mech.*, **234**, 47–69.
- Eliassen, A. and E. Raustein, 1968: A numerical integration experiment with a six-level atmospheric model with isentropic information surface. *Meteor. Annaler*, **5**, 429–449.
- Friend, A. L., D. Djuric, and K. C. Brundage, 1977: A combination of isentropic and sigma coordinates in numerical weather prediction. *Beitr. Phys. Atmosph.*, **50**, 290–295.
- Hirt, C. W., A. A. Amsden, and J. L. Cook, 1974: An arbitrary Lagrangian-Eulerian computing method for all flow speeds. *J. Comput. Phys.*, **14**, 227–253.
- Hsu, Y.-J. G. and A. Arakawa, 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Mon. Wea. Rev.*, **118**, 1933–1959.
- Johnson, D. R., 2000: Entropy, the Lorenz Energy Cycle, and Climate. In: General Circulation Model Development (D. A. Randall, ed.). *Internat. Geophys. Series*, **70**, Acad. Press, New York, 659–720.

- Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.*, **102**, 509–522.
- Kleinschmidt, E., 1950: Über Aufbau und Entstehung von Zyklonen (1. Teil). *Meteor. Rundsch.*, **3**, 1–6.
- Konor, C. S. and A. Arakawa, 1997: Design of an atmospheric model based on a generalized vertical coordinate. *Mon. Wea. Rev.*, **125**, 1649–1673.
- Lee, J.-L., R. Bleck, and A. E. MacDonald, 2009: A multi-step flux-corrected transport scheme. *Mon. Wea. Rev.*, **137**, in revision.
- Lee, J.-L. and A. E. MacDonald, 2009: A finite-volume icosahedral shallow water model on local coordinate. *Mon. Wea. Rev.*, **137**, in press.
- McDougall, T. J. and W. K. Dewar, 1998: Vertical mixing and cabbeling in layered models. *J. Phys. Oceanogr.*, **28**, 1458–1480.
- McIntyre, M. E. and T. N. Palmer, 1985: A note on the general concept for wave braking of gravity and Rossby waves. *Pure Appl. Geophys.*, **123**, 964–975.
- Phillips, N. A., 1957: A coordinate system having some special advantages for numerical forecasting. *J. Meteor.*, **14**, 184–185.
- Pierce, R., D. Johnson, F. Reames, T. Zapotocny, and B. Wolf, 1991: Numerical investigations with a hybrid isentropic-sigma model, part i: Normal mode characteristics. *J. Atmos. Sci.*, **48**, 2005–2024.
- Randall, D. A., T. D. Ringler, and R. Heikes, 2000: Global atmospheric modeling using a geodesic grid and an isentropic vertical coordinate. In: General Circulation Model Development (D. A. Randall, ed.). *Internat. Geophys. Series*, **70**, 509–536.
- Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *J. Phys. Oceanogr.*, **12**, 1154–1158.
- Rossby, C.-G., 1940: Planetary flow patterns in the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **66**, Suppl., 68–87.
- Rossby, C.-G. and Collaborators, 1937: Isentropic analysis. *Bull. Amer. Meteor. Soc.*, **18**, 201–209.
- Sadourny, R., 1975: The dynamics of finite-difference models of the shallow-water equations. *J. Atmosph. Sci.*, **32**, 680–689.
- Schaak, T. K., T. H. Zapotocny, A. J. Lenzen, and D. R. Johnson, 2004: Global climate simulation with the University of Wisconsin global hybrid isentropic model. *J. Climate*, **17**, 2998–3016.
- Takacs, L. L., 1985: A two-step scheme for the advection equation with minimized dissipation and dispersion errors. *Mon. Wea. Rev.*, **113**, 1050–1065.
- Uccellini, L. W., D. R. Johnson, and R. E. Schlesinger, 1979: An isentropic and sigma coordinate hybrid numerical model. part 1: Model development and some initial tests. *J. Atmosph. Sci.*, **36**, 390–414.
- van Leer, B., 1974: Towards the ultimate conservative difference scheme. ii. monotonicity and conservation combined in a second order scheme. *J. Comput. Phys.*, **14**, 361–370.
- Webster, S., J. Thuburn, B. J. Hoskins, and M. J. Rodwell, 1999: Further development of a hybrid-isentropic GCM. *Quart. J. Roy. Meteor. Soc.*, **125**, 2305–2331.
- Zapotocny, T., D. Johnson, and F. Reames, 1994: Development and initial test of the University of Wisconsin global isentropic-sigma model. *Mon. Wea. Rev.*, **122**, 2160–2178.
- Zapotocny, T., D. Johnson, F. Reames, R. Pierce, and B. Wolf, 1991: Numerical investigations with a hybrid isentropic-sigma model, part ii: The inclusion of moist processes. *J. Atmos. Sci.*, **48**, 2025–2043.
- Zhu, Z., J. Thuburn, B. J. Hoskins, and P. H. Haynes, 1992: A vertical finite-difference scheme based on a hybrid  $\sigma$ – $\theta$ – $p$  coordinate. *Mon. Wea. Rev.*, **120**, 851–862.